



# Secondary Datasets

Terry Watts

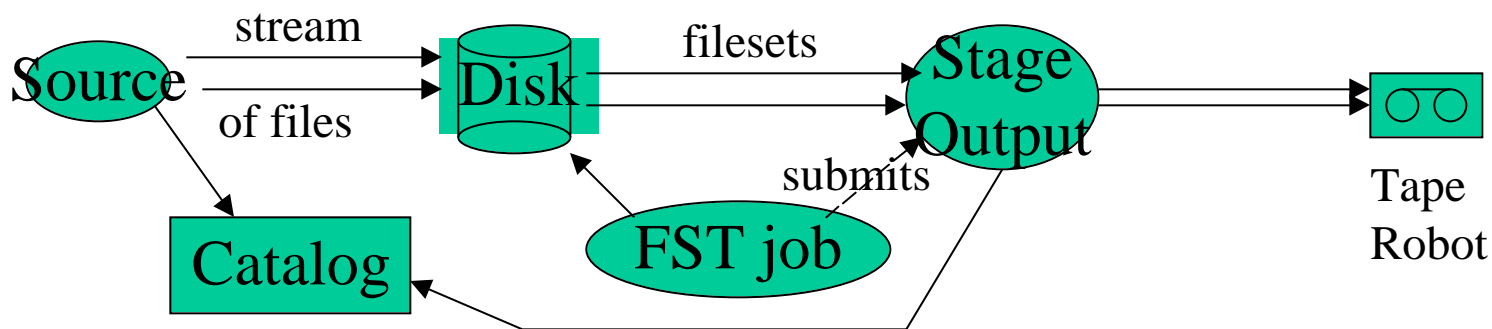
Overview in preparation for discussion of  
User and Group Datafile Catalogs by Eric  
Wicklund

DH Meeting, June 8, 2000



# Creating Dataset Files

- Basic technique (CDF Week talk):



- Source (e.g. DHOutput) makes **file**.
- **Fileset-tape(FST) daemon** forms files into filesets, and filesets into tapes.
- **Ignore today**: DIM/FST, DHOutput.



# Secondary Datasets: Other Considerations

- On Central Analysis Cluster (discuss today)
- In cdf trailers
- At remote sites
- Other types of datasets: MC, root-tuples, private, etc.
- Declaration of a dataset (see next slide)



# Declaration of a Dataset

- Before creating files, must declare a dataset.
- Propose to treat like cdf note numbers, i.e. central registry that is experiment wide (thus, applies to remote sites too). Anyone can declare a dataset.
- Dataset characterized by unique dataset\_name and unique dataset\_id. File names also must be unique.
- Dataset\_id has 6 characters: 4 are arbitrary, 2 from parent datasets (specify stream and production version).
- Not decided how to relate this registry to records in CDF2\_DATASETS in Datafile Catalog(s).